Animal Science Papers and Reports vol. 22 (2004) no. 4, 621-629
Institute of Genetics and Animal Breeding, Jastrzębiec, Poland

Is inbreeding coefficient a credible measure of autozygosity in Polish Arab horses?

Iwona Głażewska[1*], Barbara Gralak[2], Cezary Niemczewski[2]

[1] Department of Genetics and Cytology, University of Gdańsk, Kładki 24, 80-822 Gdańsk, Poland

[2] Polish Academy of Sciences Institute of Genetics and Animal Breeding, Jastrzębiec, 05-552 Wólka Kosowska, Poland

(Received August 5, 2004, accepted November 24, 2004)

The complete agreement of inbreeding or coancestry estimations based on genealogical data with those based on genetic markers in real, non-randomly mating populations has been found in some studies. However, on the basis of theoretical analysis many authors emphasise the questionable value of the inbreeding coefficient based on genealogical data as an exact measure of autozygosity (homozygosity by descent). The aim of this work was to verify whether inbreeding coefficient is an appropriate measure of autozygosity in Polish Arab horses, which are a good example of a small population strongly affected by selection. The analysis was based on pedigree and genetic markers data. The inbreeding coefficient (*F*) calculated on the basis of pedigrees was compared with heterozygosity of descendants and the similarity index (*SIM2*) based on parental genotypes. No significant correlations between *F* and heterozygosity, or *F* and *SIM2* were found. Instead, statistically significant correlations between the heterozygosity of descendants and that of their parents were found. This means that the gene frequencies and the level of heterozygosity of offspring were strictly related to the genotypes of their parents, but not to the level of inbreeding of the descendants. This statement seems to undermine the credibility of the inbreeding coefficient as a measure of homozygosity in Polish Arabs. Marker-based estimations have been found by other authors to be a useful tool for evaluating the autozygosity of a given individual when pedigree data are not available. However, the results obtained indicate the necessity of using both pedigree and marker-based estimations with the utmost care.

KEY WORDS: Arab horse / autozygosity / homozygosity / inbreeding coefficient

*e-mail: glazew@biotech.univ.gda.pl

The inbreeding coefficient is generally used to measure homozygosity of a population. Wright [1922] defined it as the correlation between uniting gametes and Malécot [1948] as the probability that two homologous alleles in an individual are identical by descent. Both definitions refer to a randomly mating, unselected population. This fact is very often omitted in publications concerning the inbreeding coefficient of domestic animals. Meanwhile, selection is of great importance in forming the gene pool of every real population, especially that of domestic animals. Analyses of theoretical models indicate the decisive significance of the influence of selection on the level of genetic variability [Lacy 1987, Curik *et al*. 2001]. Selection can have various effects leading to a decrease or increase in the initial level of heterozygosity. As is shown in theoretical simulations [Lacy 1987], directional selection leads to a decrease in variability to a level below that expected in random mating. Meanwhile, balancing selection allows for the maintenance of the initial level of variability, and even after some generations its level is higher than that under random mating.

Many authors emphasise the questionable value of the inbreeding coefficient based on genealogical data as an exact measure of autozygosity, *i.e*. homozygosity by descent. Falconer [1981] points out that the inbreeding coefficient expresses the real level of homozygosity only when factors which affect it, especially selection, do not occur. Curik *et al.* [2001] demonstrate that, depending on the direction of selection and the initial gene frequency, discrepancies are present in both directions, thus leading to the over- or underestimation of the observed autozygosity level. It occurs, therefore, that the value of the inbreeding coefficient as a measure of autozygosity is strictly dependent on the influence of selection on a population. However, as has been found in some studies, the full agreement of inbreeding or coancestry estimations based on genealogical data with those based on genetic markers in real, non-randomly mating populations may exist. Pitra *et al*. [1996] found in Asian wild horses that the inbreeding coefficient based on pedigree data was in good agreement with that based on DNA fingerprinting variability. Ellegren [1999], in his study on a captive population of grey wolves, showed that microsatellite allele sharing correlates closely with relatedness, as does individual heterozygosity with the inbreeding coefficient. Similar results, which are evidence of the significant correlation between genealogical coancestry and estimations using molecular information, were obtained in Thoroughbred horses [Cunningham *et al.* 2001] and in Iberian pigs [Toro *et al*. 2002]. These results lead to the question of whether each non-randomly mating population exhibits the same concordance between genealogical and genetic markers estimations.

The purpose of this work is to verify whether the inbreeding coefficient is a correct measure of homozygosity of Polish pure-bred Arabian horses. The Arabs, bred in Poland since the early nineteenth century, are a good example of a small, closed and highly-related population. The most important features of the population are its restricted size, the small number of founders, the high proportion of the founders present in all pedigrees, and the long-term absence of the inflow of unrelated breeding material [Głażewska and Jezierski 2004]. These factors led to the high level of relationship

of the horses. Avoiding the mating of close relatives is one of the criteria of choosing parental pairs by the breeders. Nevertheless, only one out of 10 stallions is used for breeding. Moreover, the discrepancy in breeding use of sires, which induces further increase of relationship, takes place [Głażewska 2004].

Materials and methods

Pedigree analysis

The data analysed came from 46 sires and 339 brood mares from Polish national studs, which were bred from 1987 to 1992, and their matings. These included 819 successful matings, which produced a weaned foal, and 255 unsuccessful matings ("barren mares", "abortions" and "dead foals" subgroups, the latter includes still-born, dead or eliminated foals). The pedigree data file was completed based on the Polish Arabian Stud Book (PASB), genealogical charts by Skorkowski [1960] and Rozwadowski [1972], and the publication by Kwiatkowski [1993]. The Wright [1922] inbreeding coefficients (*F*) were computed based on whole pedigrees, *i.e*. including all generations until the founders, using GENES version 11.8 software [Lacy 1998].

Genetic markers analysis

Genetic marker analysis was carried out at the Institute of Genetics and Animal Breeding, Jastrzębiec, as a routine test of parentage control. The alpha-1-beta glycoprotein (A1B), albumin (Al), carbonic anhydrase (CA), serum esterase (Es), vitamin-D-binding protein (Gc), glucose phosphate isomerase (GPI), protease inhibitor (Pi), 6-phosphogluconate dehydrogenaze (PGD), phosphoglucomutase (PGM) and transferrin (Tf) systems were analysed for the parents and their 819 offspring.

The results were analysed statistically with BIOSYS-2 software (Swofford and Selander 1981, modified by Black 1997). The observed heterozygosity (H_o), direct counting, and expected (H_e) heterozygosity, assuming Hardy-Weinberg equilibrium, within particular ranges of inbreeding of descendants were computed for offspring and their parents. The weighted frequencies of parental genotypes (*i.e*. weighted with the number of each parent's offspring) were taken into account.

To estimate differences in gene frequencies between parents and offspring, the genetic identity (*I*) was calculated according to Nei [1972] using BIOSYS-2 software.

Individual heterozygosities (H_i) as a proportion of heterozygous *loci* were calculated for the offspring. The evaluation of heterozygosity of descendants was also carried out indirectly by estimating the similarity of parental genotypes using similarity index *SIM2* [Toro *et al*. 2002]. *SIM2* was calculated as the total number of equal alleles in a pair of individuals (0, 1 or 2), divided by twice the number of *loci* analysed.

The correlations between inbreeding coefficient (*F*), heterozygosity, and similarity index (*SIM2*) were computed using MICROSOFT EXCELL 97 software.

Results and discussion

Inbreeding coefficient

The lowest and highest *F* values varied greatly in every group and ranged from 0.40% to 21.00% in the successful matings group. The *F* values in the unsuccessful matings group were as follows: 0.44%-21.28% in the "barren mare" subgroup, 0.18%-13.78% in the "abortion" subgroup, and 0.34% - 18.26% in the "dead foal" subgroup (not tabulated).

Heterozygosity

The observed and expected heterozygosities in offspring and their parents as well as the values of genetic identity between descendants and parents within every range of inbreeding are shown in Table 1. No significant correlation was found between the inbreeding coefficients and the H_o of descendants based on average values within the ranges of inbreeding (r=0.2476, P=0.3934) or on individual data (r=0.0637, P=0.0683).

Table 1. Observed (H_o) and expected (H_e) heterozygosities of descendants and their parents and genetic identity (I) across the ranges of inbreeding of descendants

Range of inbreeding (%)	Offspring		Parents		Genetic identity between offspring and parents (I)
	H_o	H_e	H_o	H_e	
0-1	0.302	0.279	0.302	0.281	1.000
>1-2	0.319	0.305	0.373	0.315	0.983
>2-3	0.320	0.308	0.314	0.312	0.998
>3-4	0.314	0.302	0.331	0.313	0.999
>4-5	0.312	0.304	0.331	0.317	1.000
>5-6	0.345	0.322	0.331	0.320	0.999
>6-7	0.342	0.320	0.351	0.323	1.000
>7-8	0.337	0.315	0.331	0.315	0.999
>8-9	0.323	0.330	0.366	0.336	0.999
>9-10	0.323	0.329	0.375	0.338	0.999
>10-11	0.341	0.327	0.338	0.341	0.994
>11-12	0.359	0.381	0.391	0.370	0.997
>12-14	0.362	0.379	0.362	0.337	0.995
>14	0.289	0.291	0.322	0.308	0.997
Total	0.328	0.320	0.340	0.324	1.000
Correlation[1]	0.2476				
P	0.3934				
Correlation[2]	0.0637				
P	0.0683				

The mean observed heterozygosity was higher than that expected in both descendants and parents, with the latter being more heterozygous. There was also a high level of genetic identity (*I*) between descendants and parents irrespective of the inbreeding level (*F*) of the former. The lowest *I* value (0.983) was estimated for the >1%-2% inbreeding range, while it varied from 0.994 to 1.0 for the remaining inbreeding ranges. The *I* value calculated for the total of the ranges for both descendants and parents was 1.0. This means that the difference in the heterozygosities between descendants (H_o=0.328) and parents (H_o=0.340) was not accompanied by significant differences in gene frequencies, which could be reflected in the decrease of genetic identity between the two groups.

Significant correlations were found between the H_o of descendants and that of parents, between the H_o of descendants and the H_e of parents and also between the H_e of descendants and the H_e of parents (Tab. 2).

Table 2. Correlations (*r*) between observed (H_o) and expected (H_e) heterozygosities of descendants and their parents, computed on the basis of H_o and H_e given in Table 1

Group		Offspring	Parents	
		H_e	H_o	H_e
Offspring	H_o	*r*=0.8649 P=0.0010	*r*=0.5398 P=0.0463	*r*=0.7105 P=0.0044
	H_e		*r*=0.7324 P=0.0029	*r*=0.8769 P=0.0000
Parents	H_o			*r*=0.8146 P=0.0004

Similarity index

In the second stage of analysis, the homozygosity of matings was estimated indirectly by calculating *SIM2*. Firstly, it was checked whether *SIM2* is an appropriate measure for evaluating the heterozygosity of matings by comparing individual heterozygosities of offspring (H_i) and the values of *SIM2* based on their parents' genotypes. A significant, negative correlation between H_i and *SIM2* was found (*r*=0.4075; P=0.0000) in the 819 successful matings which confirms the usefulness of *SIM2* for present purposes.

Table 3 shows the *SIM2* values referred to the level of inbreeding within each type of mating. In no cases were significant correlations found between *F* and *SIM2.* Calculations were also done separately for each gene (data not shown), but significant relationships were not found in any case.

The level of autozygosity of the population was evaluated using two parameters. The first was heterozygosity (H_o) calculated on the basis of the genotypes of descendants born in the analysed period, and the second was the similarity index (*SIM2*) com-

Table 3. Mean inbreeding coefficients (*F*) and similarity indexes (*SIM2*) in the groups differing in mating results

Range of inbreeding (%)	Unsuccessful matings												Successful m (offsprin	
	barren mares			abortions			dead foals			total				
	n	*F* (%)	*SIM2*	n	*F* (%)	*SIM2*	n	*F* (%)	*SIM2*	n	*F* (%)	*SIM2*	n	*F* (%)
0-1	2	0.45	0.7750	4	0.53	0.7500	8	0.48	0.7688	14	0.49	0.7643	43	0.56
>1-2	8	1.36	0.7563	–	–	–	2	1.55	0.7000	10	1.40	0.7450	14	1.37
>2-3	5	2.44	0.6900	–	–	–	1	2.63	0.7500	6	2.47	0.7000	45	2.75
>3-4	13	3.56	0.7269	4	3.68	0.7375	11	3.68	0.7182	28	3.62	0.7250	97	3.60
>4-5	20	4.42	0.7600	6	4.51	0.7000	10	4.42	0.7500	36	4.43	0.7472	116	4.52
>5-6	22	5.40	0.7114	4	5.42	0.7000	8	5.39	0.7375	34	5.40	0.7162	131	5.51
>6-7	22	6.58	0.7273	2	6.26	0.7750	13	6.53	0.7346	37	6.55	0.7324	130	6.52
>7-8	15	7.39	0.7800	4	7.48	0.8250	10	7.39	0.7300	29	7.40	0.7690	86	7.45
>8-9	12	8.54	0.7250	–	–	–	3	8.35	0.7167	15	8.50	0.7233	48	8.52
>9-10	6	9.55	0.7667	2	9.48	0.6750	3	9.50	0.7000	11	9.52	0.7318	35	9.52
>10-11	7	10.79	0.7143	3	10.28	0.7500	5	10.72	0.7100	15	10.66	0.7200	24	10.48
>11-12	3	11.71	0,7833	1	11.59	0.8000	3	11.27	0.7333	7	11.51	0.7643	16	11.36
>12-14	4	13.18	0.7875	2	13.25	0.7750	–	–	–	6	13.18	0.8233	16	12.79
>14	6	16.93	0.7250	–	–	–	1	18.26	0.9000	7	17.12	0.7500	18	16.28
Total	145	6.60	0.7393	32	6.12	0.7422	78	5.84	0.7353	255	6.31	0.7384	819	6.01
Correlation*		0.0371			0.1722			-0.0005			0.0431			0.0291
P		0.6575			0.3460			0.9966			0.4934			0.4055

*Correlation coefficient (*r*) between *F* and *SIM2*, computed on the basis of individual data.

puted on the basis of parental genotypes. The agreement was expected between the inbreeding coefficient (*F*) and the homozygosity of an individual (H_i) as manifested in the significant negative correlation between *F* and *H* and/or significant positive correlation between *F* and *SIM2*. Instead, irrespective of the measure used - either heterozygosity or *SIM2* – no significant correlations between the *F* of a given individual and its homozygosity were found. The lack of the expected correlation between *F* and the heterozygosity of weaned foals might suggest that a considerable number of matings that lead to the higher autozygosity of descendants culminate in barrenness or abortion or in the death of the foal prior to the age at which genetic markers analysis is carried out. For obvious reasons, it is impossible to investigate genetic markers in most cases of unsuccessful mating. Therefore, the heterozygosity of potential offspring was estimated indirectly by evaluating *SIM2*. However, the negative results for both successful and unsuccessful matings were also obtained with this method.

The analysis confirmed the very high level of genetic identity (*I*) between descendants and their parents, independent of the level of inbreeding. It was also confirmed that the heterozygosity of descendants was significantly correlated with the heterozygosity of their parents, involving both observed and expected heterozygosities. These results indicate that gene frequencies and heterozygosity in offspring are strictly related to the genotypes of their parents, but not to the inbreeding level of descendants.

It is noteworthy that in both descendants and parents the mean H_o was higher than the H_e. The H_o predominated over the H_e in 9 out of 14 ranges of inbreeding in the descendants and in 13 out of 14 ranges in the parents (Tab. 1) indicating that the partial elimination of homozygotes might occur at the prenatal stage as a result of natural balancing selection. In addition to natural selection, which is expressed through the higher mortality of homozygous individuals, selection based on breeding preferences (breeder decisions and fertility differences) occurs at further stages and favours more heterozygous horses. As is demonstrated by Głażewska [2004], the statistically significant preference of less inbred Arab stallions while being selected for the breeding stock has been noted in the space of fifty years of Polish breeding. Also, the level of allozyme heterozygosity of the female offspring chosen for breeding is higher than that of non-breeding mares (authors' unpublished data).

The fundamental goal of Polish Arab breeding is to maintain the true desert type of horse, which embodies great beauty, soundness and racing stamina. The additional selection criterion for sires is their perceptibly higher vitality. This set of features seems to be subject to balancing rather than directional selection, the latter of which commonly appears in animal breeding. It may be supposed that selection, which leads to the partial elimination of homozygotes in Polish Arabs, occurs in every succeeding generation. Although the selection pressure in every generation is weak, it seems to be of basic importance for the genetic variability of the analysed population because of its long-term impact [Falconer 1981]. It should be pointed out that the inbreeding coefficients used in this study were calculated on the basis of whole pedigrees (even back to the 20th generation) so they reflect long-term selection pressure. As is shown

in theoretical simulations [Lacy 1987, Curik *et al*. 2001], balancing selection leads to higher heterozygosity than is expected in a randomly mating population. This then leads to the discrepancy between the real level of homozygosity and that expected on the basis of the inbreeding coefficient. The results of the present study confirm the above theoretical expectations.

Unlike Polish Arabs, statistically significant correlations between pedigree and genetic markers estimations were found in Asian wild horses [Pitra *et al*. 1996], grey wolves [Ellegren 1999], Iberian pigs [Toro *et al*. 2002] and Thoroughbred horses [Cunningham *et al*. 2001]. Selection performed within these populations differs from that applied in Polish Arabs. Asian wild horses, grey wolves and Iberian pigs are examples of endangered populations subjected to conservation programmes whose main goal is to maintain an unchanged level of genetic variability. The basic criterion of selection in Thoroughbreds is racing performance, which indicates directional selection in the breed. This type of selection leads to higher homozygosity than that expected in a population under random mating. The results presented in this work and those from the articles cited seem to confirm that the reliability of the inbreeding coefficient as a measure of autozygosity is strictly conditioned by the influence of selection on a given population.

Ellegren [1999] found marker estimations to be a useful tool for evaluating the autozygosity of an individual when pedigree data are not available. However, the results of the present study indicate that pedigree- or marker-based estimations should be used with the utmost care. Such caution is also motivated by the results of theoretical simulations [Baumung and Sölkner 2003], which clearly show that the credibility of the estimation is dependent on the number of animals subjected to analysis, the length, completeness and correctness of pedigrees and the number of genetic markers used. All these factors can have significant influence on the findings.

REFERENCES

1. BAUMUNG R., SÖLKNER J., 2003 – Pedigree and marker information requirements to monitor genetic variability. ***Genetics, Selection, Evolution*** 35, 369-383.
2. CUNNINGHAM E.P., DOOLEY J.J., SPLAN R.K., BRADLEY D.G., 2001 – Microsatellite diversity, pedigree relatedness and the contributions of founder lineages to thoroughbred horses. ***Animal Genetics*** 32, 360-364.
3. CURIK I., SÖLKNER J., STIPIC N., 2001 – Effects of models with finite loci, selection, dominance, epistasis and linkage on inbreeding coefficients based on pedigree and genotypic information. ***Journal of Animal Breeding and Genetics*** 119, 101-115.
4. ELLEGREN H., 1999 – Inbreeding and relatedness in Scandinavian grey wolves *Canis lupus*. ***Hereditas*** 130, 239-244.
5. FALCONER D.S., 1981 – Introduction to Quantitative Genetics. Longman, New York.
6. GŁAŻEWSKA I., 2004 – Mating and selection in national Arabian horse breeding: analysis of inbreeding coefficients. ***Electronic Journal of Polish Agriculture Universities, Animal Husbandry*** 1, 7. Available by Internet www.ejpau.media.pl
7. GŁAŻEWSKA I., JEZIERSKI T., 2004 – Pedigree analysis of Polish Arabian horses based on founder contributions. *Livestock Production Science* 90, 293-298.

8. KWIATKOWSKI W., 1993 – A guide to the complete pedigrees of Arabian horses from Poland. Published by Kawalkada, Brwinów.
9. LACY R.C., 1987 – Loss of genetic diversity from managed populations: Interacting effects of drift, mutation, immigration, selection, and population subdivision. ***Conservation Biology*** 1 (2), 143-158.
10. LACY R.C., 1998 – GENES. Version 11.8. Software package for genetic analysis of studbook data. Available by: www.vortex9.org/genes.html
11. MALÉCOT G., 1948 – Les Mathématiques de L'Hérédité. Masson et Cie, Paris.
12. NEI M., 1972 – Genetic distance between populations. ***American Naturalist*** 106, 283-291.
13. PITRA C., CURSON A., NÜRNBERG P., KRAWCZAK M., BROWN S., 1996 – An assessment of inbreeding in Asian wild horse (*Equus przewalski* Poliakow 1881) population using DNA fingerprinting. ***Archiv Fuer Tierzucht*** 39, 589-596.
14. ROZWADOWSKI Z., 1972 – 50 years of breeding pure bred Arabian Horses in Poland in their genealogical charts 1918-1968. PWRiL, Warszawa.
15. SKORKOWSKI E., 1960 – Tablice genealogiczne polskich koni arabskich czystej krwi (Genealogical Tables of Polish Arab Horses). In Polish. Instytut Zootechniki, Kraków.
16. TORO M., BARRAGÁN C., ÓVILO C., RODRIGAÑEZ J., RODRIGUEZ C., SILIÓ L, 2002 – Estimation of coancestry in Iberian pigs using molecular markers. ***Conservation Genetics*** 3, 309-320.
17. WRIGHT S., 1922 – Coefficients of inbreeding and relationship. ***American Naturalist*** 645, 330-338.

Iwona Głażewska, Barbara Gralak, Cezary Niemczewski

Czy współczynnik inbredu jest wiarygodną miarą autozygotyczności polskich koni arabskich?

S t r e s z c z e n i e

Wielu autorów opracowań teoretycznych zwraca uwagę na wątpliwą wartość współczynnika inbredu jako ścisłej miary homozygotyczności „przez dziedziczenie" (autozygotyczności). Jednakże w pewnych publikacjach potwierdzono pełną zgodność między poziomami inbredu i pokrewieństwa oszacowanymi na podstawie rodowodów a oszacowanymi na podstawie markerów genetycznych. Celem pracy było sprawdzenie, czy współczynnik inbredu jest właściwą miarą autozygotyczności polskich koni arabskich. W analizie wykorzystano dane rodowodowe oraz genetyczne. Wartości współczynnika inbredu (*F*), obliczonego na podstawie rodowodów, porównano z heterozygotycznością potomków oraz wartościami wskaźnika podobieństwa *SIM2*, obliczonego na podstawie genotypów rodziców. Stwierdzono brak korelacji między porównywanymi parametrami. Zaobserwowano natomiast istotne zależności między heterozygotycznością potomków a heterozygotycznością ich rodziców. Oznacza to, że frekwencja genów i heterozygotyczność potomków są ściśle powiązane z genotypami rodziców, a nie z poziomem inbredu, obliczonym na podstawie rodowodów. Tym samym uzyskane wyniki wydają się podważać wiarygodność współczynnika inbredu jako ścisłej miary homozygotyczności polskich koni arabskich.

Szacunki oparte na analizie markerów uznawane są przez różnych autorów za użyteczne narzędzie oceny autozygotyczności danego osobnika w sytuacji, gdy brakuje danych rodowodowych. Jednakże przedstawione wyniki wskazują, że przy wnioskowaniu o autozygotyczności osobnika, zarówno na podstawie wyników analizy genetycznej jak i analizy rodowodowej, należy zachować daleko posuniętą ostrożność.